

Title Page

The implications of programme assessment patterns for student learning

First Author/Corresponding author

Professor Tansy Jessop
Solent Learning and Teaching Institute
Southampton Solent University
East Park Terrace
Southampton
SO14 0YN
Phone: 02382016982
Tansy.Jessop@solent.ac.uk

Second Author

Dr Carmen Tomas
Teaching and Learning Directorate
University of Nottingham
Nottingham
NG7 2RD
Phone: 0115 7484772
Carmen.Tomas@nottingham.ac.uk

Biographical note:

Professor Tansy Jessop has led the TESTA project for seven years. She has previously published on disciplinary assessment patterns and the implications of modular degrees on assessment patterns. She is Professor of Research Informed Teaching at the Solent Learning and Teaching Institute, Southampton Solent University.

Dr Carmen Tomas is Assessment Adviser in the University of Nottingham's Teaching and Learning Directorate. She has adapted the TESTA methodology to help improve assessment design on various programme at Nottingham.

The implications of programme assessment patterns for student learning

Tansy Jessop

Southampton Solent University

Carmen Tomas

University of Nottingham

Abstract

Evidence from 73 programmes in 14 UK universities sheds light on the typical student experience of assessment over a three year undergraduate degree. A previous small-scale study in three universities characterised programme assessment environments using a similar method. The current study analyses data about assessment load parameters using descriptive statistical methods, drawing on a large sample in a wider range of universities than the original study. Findings demonstrate a wide range of practice across programmes: from 12 summative assessments on one programme to 227 on another; from 87 per cent by examination to none on others. While variations cast doubt on the comparability of UK degrees, programme assessment patterns are complex. Further analysis distinguishes common assessment patterns across the sample. Typically students encounter eight times as much summative as formative assessment, a dozen different types of assessment; more than three quarters by coursework. The presence of high summative and low formative assessment diets is likely to compound students' grade-orientation, reinforcing narrow and instrumental approaches to learning. High varieties of assessment are probable contributors to student confusion about goals and standards. Making systematic headway to improve student learning from assessment requires a programmatic and evidence-led approach to design, characterised by dialogue and social practice. (200 words)

Introduction

Recent studies question the depth of student learning from assessment on modular degree programmes over short semesters, in the UK and elsewhere, for example, India, Australia and New Zealand (Knight and Yorke 2003; Jessop, El Hakim and Gibbs 2014; Harland, McLean, Wass, Miller, and Sim 2015). Modular degrees have afforded choice, flexibility and the opportunity for students to curate their own curriculum. These characteristics are likely to have improved students' motivation to study. But the benefits of choice may be outweighed by the negative effects of compartmentalised modules, which disrupt coherence by fragmenting learning across whole programmes of study. Assessment is a key component in curriculum design, and is often the place where a lack of articulation across modules is most evident, with programmes designed as "islands apart from the bodies of knowledge and practices from which they are generated and on which they focus" (Boud and Falchikov 2006, 406). It is common practice for programme leaders to compile degree programmes by collecting individual modules from team members, in an assembly of parts where the articulation between the parts and the whole is weak (Jessop, El Hakim and Gibbs 2014). From a student perspective, this leads to a fragmented and disconnected learning experience.

Poorly designed assessment has significant bearing on learning, because students direct most of their time, effort and attention at undertaking assessment (Gibbs and Simpson 2004). Curriculum design which ignores the significance of assessment in prompting student learning, fails to acknowledge that “assessment is the senior partner in learning and teaching” (Biggs and Tang 2011, 221). The predominance of assessment in students’ minds shapes their relationship with the curriculum: “from our students’ point of view, assessment always defines the actual curriculum” (Ramsden 2003). Assessment is the engine which powers student engagement and effort.

Student ratings on the National Students’ Survey (NSS) indicate that assessment and feedback is the weakest link in their university experience. Over a ten year period, NSS scores for assessment and feedback have persistently been at least 10 per cent lower than those about the quality of teaching provision and students’ overall satisfaction (Williams and Kane, 2009). Explanations for low student satisfaction scores are complex. They include: the impact of large class sizes on the quality and timing of feedback (Calvo, Howard and Markauskaite 2010; Gibbs, 2013); imagined and real variations in the quality and quantity of feedback (Weaver 2006; Price, Handley, Millar and O’Donovan 2010); student expectations that feedback will provide a pedagogic relationship in an era of mass higher education, and their disappointment with it (Nicol 2010); reliance on written criteria to deliver clarity about standards, and the lack of opportunity for students to internalise standards (Sadler 2007; Nicol and McFarlane-Dick 2006). This article explores whether these explanations are partial reflections of a more systemic cause of students’ dissatisfaction, namely, the lack of overarching assessment and feedback design across whole programmes of study.

Large-scale evidence from TESTA tells a powerful story about disconnected and under-theorised assessment design (Jessop, El Hakim and Gibbs 2014). TESTA (Transforming the Experience of Students through Assessment) is a programme-focused research approach which has been used to transform assessment practice on hundreds of degrees in more than 50 universities in the UK, and in universities in India and Australia. The central argument in this article is that persistent dissatisfaction with assessment and feedback in the NSS is largely a consequence of fragmented design. Similar arguments about the need for more holistic assessment and feedback design in a context of fractured modular systems, have been advanced in the literature (Bloxham and Boyd 2007; Boud and Molloy 2013). This article brings to the debate robust empirical evidence about the effects of fractured design.

Addressing these important issues about assessment design and student experience requires a deeper understanding of current practice and typical assessment loads. There is only one previous study that attempts to demonstrate the assessment load at programme level and from a student perspective. In a previous study characterising assessment environments using the TESTA audit methodology, coding boundaries were set based on data from nine programmes at three universities (Gibbs and Dunbar-Goddet 2009). This early work discussed disparities across one post-92, one Oxbridge and a pre-92 ‘red brick’ institution, using coding boundaries of High, Medium and Low. These categories were said to distinguish different assessment environments by discipline and university-type. The authors described the process of devising coding boundaries as ‘arbitrary except that they succeed in distinguishing between programmes’ (*Ibid*, 486). The authors called for a larger-scale study, given the limitations of the size of their sample (Gibbs and Dunbar-Goddet 2009, 488). The current study seeks to build from this small scale study to expand and explore patterns of assessment and feedback, providing insights into the typical student experience. The paper

represents a contribution in terms of the larger sample and more replicable analysis deployed using explicit descriptive statistics.

Method

TESTA triangulates programme data from three sources: an audit based on official programme documents and the programme leader's interpretation of these; quantitative data from final year students who complete the Assessment Experience Questionnaire; and focus group data reflecting final year perspectives on the programme under study. This article focuses on the audit data, although the interpretation draws on wider findings from the TESTA process (Jessop, El Hakim and Gibbs 2014; Jessop and Maleckar 2014).

The audit is designed to elicit frequency counts of aspects of assessment and feedback on whole degree programmes. It consists of a two hour interview conversation with the programme leader over the validated degree documents. These documents capture the main features of the planned curriculum, for example, the number and types of summative assessment, and the balance of examinations and coursework. Definitive documents are more opaque about formative assessment tasks, as many of these are discretionary, both for lecturers and students (Jessop, McNab and Gubby 2012). Conversations with the programme leader give a clearer picture of formative assessment loads. Further data is collected through sampling marked scripts, and a word count of written words of feedback in summary comments and in the margins. The uniqueness of the TESTA audit lies in its capacity to capture assessment data across a whole programme rather than on a single module.

The TESTA audit involves collecting data on the following dimensions of practice:

- Number of summative tasks;
- Number of formative-only tasks;
- Varieties of assessment;
- Proportion of examinations;
- Amount of written feedback;

The TESTA methodology defines summative assessment as measuring student achievement, in contrast to formative which is designed to foster learning. The key elements of formative assessment within the TESTA definition are that: it needs to be done by all students, it elicits feedback and it does *not* count towards the degree. While the literature on learning-oriented assessment (Carless 2007, 2015) proposes that summative assessment can be designed to help students learn, modular degree systems obstruct learning from summative assessment through disconnected and compartmentalised assessment and feedback design (Jessop, El hakim and Gibbs 2014, Jessop and Maleckar 2014; Harland, McLean, Wass, Miller, and Sim 2015).

The sample

The sample in this study consists of 73 programmes, spread virtually evenly across research intensive (35) and teaching-oriented universities (38). The sample of disciplines is skewed more towards Humanities, Arts and Social Sciences than STEM subjects (Science, Technology, Engineering and Mathematics). 52 out of 73 programmes are from more interpretive and divergent 'soft' disciplines and 21 programmes represent 'hard' STEM subjects (Biglan 1973a; 1973b; Becher and Trowler

2001). The ratio of soft to hard disciplines is about 2.5 to 1. Previously, a study of 18 programmes in eight universities based on the TESTA methodology showed that there were statistically significant differences in features of assessment and feedback in different discipline groups (Jessop and Maleckar 2014). Although disciplinary analysis is not the focus of this paper, the authors acknowledge that the data may lean more towards representing assessment and feedback patterns in 'soft' disciplines. A further paper which focuses on exploring these disciplinary differences is being prepared.

Data analysis

The researchers used descriptive statistical analysis to determine coding values corresponding to Low, Medium or High assessment patterns on the 73 programmes in the sample. Descriptive statistics for non-normally distributed data have been chosen for their greater resistance to the presence of outliers in the data. The coding boundaries that determine Low, Medium and High are derived using quartiles. Quartiles define the position of a given value when these are placed in order of increasing size. Low, Medium and High boundaries are defined using the first, second and third quartiles. In summary the boundaries are:

- Low – below the first quartile value represents the range in the first quarter
- Medium – above the first quartile threshold and below the third quartile threshold
- High – above the third quartile representing the top quarter

Using these values also serves to display the interquartile range between low and high boundaries to provide a measure of the dispersion of the data.

Findings

The current study evidences key differences in programme assessment and feedback environments compared to the original study. These differences occur in the analysis of ranges of assessment features; and within coding categories of High, Medium and Low boundaries of practice. Finally, the findings section compares medium categories across the two studies, highlighting current assessment patterns. Current patterns form the basis for exploring implications for student learning.

A comparison of ranges related to assessment characteristics across the two studies provides a tentative picture, given the likelihood of outliers. Table 1 illustrates ranges from Study 1 (2009) and Study 2 (the current study). Study 2 has a much wider range of: summative assessment, return times of feedback, and written feedback than Study 1. Study 1 has wider variations of oral feedback than Study 2. Plausible deductions from the data are that: a) Oxbridge's tutorial system has contributed to the high end of the oral feedback range in Study 1; b) higher incidence of large classes in Study 2 may contribute to extended length of feedback return times and the absence of one day return of feedback.

Table 1: Comparison of assessment characteristic ranges using the same methodology

Characteristics	Study 1 (2009) n=3 universities n= 9 programmes	Study 2 (2016) n=14 universities n=73 programmes
-----------------	---	--

Number of Summative Assessments	11 - 95	12 - 227
Number of Formative Assessments	2 - 134	0 - 116
Varieties of assessment	2 - 18	5 - 19
Proportion of tasks by exam	17% - 100%	0% - 87%
Time to return marks & feedback	1 – 28 days	10 – 42 days
Volume of oral feedback	180 – 4,080 minutes	40 -1,800 minutes
Volume of written feedback	2,700 – 10,350	950- 22,000

While the ranges provide cause for concern about the comparability of student experience on different programmes in various universities, no conclusions can be drawn from them about a typical student experience, given the presence of statistical outliers. It requires a more sophisticated method of establishing coding boundaries to distinguish accurately the typical student experience of assessment and feedback.

The original research on nine programmes in three contrasting universities focused on ranges and the categorisation of these ranges into Low, Medium and High. Because the sample was small, the data from the Oxbridge university is bound to have skewed the categorisation. At Oxbridge, students study one discipline in depth, with very little choice in the curriculum, using summative assessment to measure achievement either at year levels or the degree programme level. Formative assessment is central to Oxbridge's tutorial system, as are final examinations which integrate learning from across the whole degree. Essays and examinations are its main summative currency, which contrasts with the emphasis in post-1992 universities, where there is greater variety and higher proportion of coursework assessment (Gibbs and Dunbar-Goddet 2009). Given the characteristics of Oxbridge and the small sample, it is not surprising that the coding shows higher formative, lower summative, lower variety, and a higher proportion of examinations than is typical across the sector.

The original categorisation was until now was the only published set of assessment parameters at the disposal of the sector. Table 2 shows the categorisation by Low, Medium and High coding boundaries from the study, with grey shading indicating typical Oxbridge extremes.

Table 2: Definitions of High, Medium and Low (Gibbs and Dunbar-Goddet 2009).

Characteristic	Low	Medium	High
Volume of summative assessment	Below 15	15-40	More than 40
Volume of formative only	Below 15	15-40	More than 40
per cent of marks from examinations	Below 40%	40-70%	70%

Variety of assessment methods	1-3 methods	4-6 methods	More than 6 methods
Written feedback	Less than 3000 words	3000-6000 words	More than 6000 words

Defining categories in a larger sample of more representative universities, using interquartile analysis, paints a very different picture of UK assessment parameters to the original study. The medium category, representing 50 per cent of the sample, has a narrower range than the earlier study in all categories except varieties of assessment. This shows that the middle 50 per cent of students are experiencing broadly similar assessment patterns, aside from their experience of formative assessment. Typically, most have a higher number of summative assessments (40-48) than the 15-40 originally characterised; a lower band of formative (5-19) than the 15-40 in the original sample. The proportion of exams is lower and narrower in range (22 per cent -31 per cent) than the original 40 per cent -70 per cent. Students experience more than double the varieties of assessment (11-15) than in the original sample (4-6). Finally, students typically receive more written feedback (6,000 -7,600) than the 3,000-6,000 originally found. Table 3 portrays the typical assessment pattern of students in the Medium category.

Table 3: Definitions of High, Medium and Low on sample of 73 programmes, using Interquartile Ranges

Characteristic	Low	Medium	High
Volume of summative assessment	Below 33	40-48	More than 48
Volume of formative only	Below 1	5-19	More than 19
% of marks from examinations	Below 11%	22-31%	More than 31%
Variety of assessment methods	Below 8	11-15	More than 15
Written feedback in words ¹	Less than 3,800	6,000-7,600	More than 7,600

Table 4 compares Gibbs and Dunbar-Goddet's (2009) medium boundaries with current study, which gives a more accurate portrayal of the typical student experience of assessment and feedback on a whole degree programme.

Table 4: Comparison of 'Medium' Assessment and Feedback characteristics

Characteristic	Medium 2009 study n=9	Medium 2016 study n=73 programmes
Volume of summative assessment	15-40	40-48

¹ The original figures of 3,781; 5,982-7,639; 7,639 have been rounded up or down to the nearest hundred for neatness sake.

Volume of formative only	15-40	5-19
% of marks from examinations	40-70%	22-31%
Variety of assessment methods	4-6	11-15
Written feedback in words	3000-6000	6,000-7,600

The current study provides a reasonable degree of confidence for establishing the parameters of a typical experience of assessment and feedback across a programme. In the discussion which follows, the authors explore implications for student learning.

Implications of typical assessment patterns for student learning

The discussion explores three consequential implications for student learning. The first is the persistence of variations in assessment and feedback across different programmes of study; the second the high volumes of summative assessment compared to formative; and thirdly, the high variety of assessment types students encounter. The discussion poses questions about whether compartmentalised assessment design is producing compartmentalised learning.

a. Wide variations

The persistence of wide variations in assessment and feedback is problematic, even when taking into consideration the presence of statistical outliers. It has long been speculated that the amount of summative assessment on degree programmes differs markedly: “there is considerable variation in the amount of work that is assessed for grading or classification” (Knight, 2002, 110). Empirical evidence about variations of assessment load on UK degrees has been more elusive, with the only attempt to classify students’ programme-wide experience arising in Gibbs and Dunbar-Goddet’s research (2009). The consequence of high assessment loads for students is the tendency to adopt a surface approach to learning (Knight 2002; Lizzio, Wilson and Simons, 2002).

The literature indicates plausible explanations for some extent of variation, for example different disciplinary practices. Different “ways of thinking and practicing” in disciplines may be reflected in how students are assessed (Hounsell and Anderson 2009). In professional contexts, disciplines have “signature pedagogies” which link to the particular ways in which different professions enact their knowledge (Shulman 2005). Distinctive disciplinary and professional patterns of assessment and feedback have been shown in previous research (Warren-Piper, Nulty and O’Grady 1996; Jessop and Maleckar 2014). However, variations may also be an accidental consequence of modular design. It is not unusual for programmes to be assembled module-by-module according to the content being covered, without an overarching eye on the entire assessment pattern. Module leaders may design assessment tasks independently from each other without any necessary connection to the wider programme (Jessop, McNab and Gubby, 2012).

Wide variations imply that elements of the assessment environment are invisible and unregulated. Surprisingly, variations exist in spite of the ever-present quality assurance apparatus, both

institutional and external (Knight 2002). Institutional quality assurance processes tend to focus on scrutinising assessment design at a granular and modular level, for example, word lengths in relation to levels of study; the coherence of assessment on individual modules; the comparability of assessment in relation to levels of study; and the alignment of particular assessment tasks with learning outcomes. Optional modules obscure a holistic programme view. Given the complexity of modular choices within degree programmes, it is difficult for those assuring the quality to get a full picture of the assessment pattern *across a whole programme*.

b. High summative: low and varied formative

The data shows that most students experience a high incidence of summative assessment combined with a low and highly varied incidence of formative tasks. Students typically experience a ratio of 1:8 of formative to summative tasks over the course of a degree programme. Formative tasks and feedback are known to be powerful drivers of student learning (Black and Wiliam 1998; Hattie 2007). While summative assessment can help students to learn (Carless 2007) its principal purpose is to measure student achievement. It orients students towards grading and towards a narrow interpretation of precise criteria and specified learning outcomes, in a way that often leads to an instrumental behaviour, and technical prowess at meeting the criteria. Critiques of high summative assessment diets suggest that they foster grade-orientation at the expense of learning (Becker, Geer and Hughes 1968; Jessop, El Hakim and Gibbs 2014; Harland, McLean, Wass, Miller, and Sim 2015). High stakes summative assessment often works against creativity, complex learning and risk-taking because students are focused on addressing criteria and achieving good grades which count towards the final degree award. Arguably ‘assessment *as* learning’ comes to replace ‘assessment *of* learning’ in a context of tightly specified criteria, as the achievement of grades may be at the expense of a more complex understanding (Sadler 2007; Torrance 2007). The apparatus of summative assessment sets it up to foster performativity.

Summative assessment has become one of the main academic drivers to activate student effort and control study behaviour. Assessment is “the principal mechanism whereby staff exercise power and control over students” (Boud 1999, 418). A study of student learning responses to a high summative assessment diet in a university in New Zealand indicated that summative assessment was used as a ‘pedagogy of control’. It prompted students to prioritise small, frequent, and often trivial assessment tasks at the expense of meaningful engagement in teaching and learning, setting up a treadmill of effort for students to accumulate a succession of small grades adding up to a degree (Harland, McLean, Wass, Miller, and Sim 2015). There was competition between concurrent modules for student time and attention, leading to an ‘assessment arms race’, and preventing lecturers climbing down from an extremely high number of assessment tasks. There are sound pedagogical reasons for encouraging students to spend more ‘time-on-task’ as this is an important factor in developing student understanding (Gibbs and Simpson 2004), but high summative diets without accompanying formative tasks induce instrumental study behaviour.

Reasons for the relative absence of formative assessment are complex. One reason is that formative tasks compete for time with summative ones, in a context of relatively brief semesters. In the absence of a programmatic decision to substitute summative with formative assessments, lecturers on single modules who adopt formative approaches struggle to gain student buy-in. Students focus

more attention on what counts towards their degree when submission deadlines collide (Jessop, El Hakim and Gibbs 2014). A second and less well-explored reason is the difficulty academics face in designing compelling formative tasks which engender enthusiasm among students, in a context of grade-orientation, and when students are pressurised by high volumes of summative assessment (Jessop, McNab and Gubby 2012). A third reason is that academics are reluctant to reduce summative assessments in a climate of appeals and complaints, particularly when assessing students from a wider range of backgrounds and abilities than ever before. Logic dictates that the fewer the summative, the higher the stakes. Students collude with academics in a 'disengagement compact', which is code for "I'll leave you alone if you'll leave me alone" (Kuh, 2003, 28). This promotes a cautious culture of business-as-usual. In most institutions, internal quality assurance departments prefer a quiet life and reinforce the status quo. In the end, the challenge of motivating students to undertake formative tasks surmounts the potential value of those tasks. The idea that well-executed formative assessment could revolutionise student learning has not yet taken hold.

c. Varieties of assessment types interfere with clarity of understanding

New data about assessment varieties contrasts sharply with the previous study which provided a modest picture of assessment varieties in the medium category (4 - 6 types). The current study depicts a medium category of between 11 and 15 varieties of assessment within students' 40 to 48 summative assessment tasks. Examples of varieties include examinations, essays, reflective journals, posters, presentations, critical commentaries, laboratory reports, films, portfolios, blogs, viva voce examinations, objective structured clinical examinations (OSCEs), projects, annotated bibliographies, and research dissertations. Extrapolating the number of varieties in the medium category, if each occurred equally often this would mean that students were likely to experience each type of assessment approximately four times over three years. Even if this were the case, it would imply a steep learning curve for students mastering complex assessment types, particularly if these were not carefully sequenced through the degree. However, in gathering audit data, two types of assessment predominate (essays and examinations), with creative methods occurring randomly and infrequently through the degree, often as one-off occurrences.

The introduction of more experimental methods of assessment is a consequence of a design approach where modular autonomy is stronger than programme coherence. Random and episodic creativity in assessment is often disconnected from the broader programme philosophy, causing students to struggle to understand its relationship to the bigger picture. Evidence from TESTA suggests that a high incidence of varieties may confuse students, and impair their grasp of goals and standards (Jessop, El Hakim and Gibbs 2014). This is particularly the case in an environment of low and weakly conceived formative assessment. Students benefit from low-risk formative tasks, with feedback which helps them to fine-tune and clarify understanding, especially in encounters with new types of assessment (Sadler 1989, Boud 2000). TESTA statistical data has corroborated a lack of clarity about goals and standards among students, but has not yet provided a full explanation of reasons for it (Jessop, El Hakim and Gibbs 2014). This study suggests that one explanation may lie in the high variety of assessment which students encounter.

Conclusion

This article adds to the literature by providing a clear indication of assessment and feedback

patterns which students commonly encounter on their degree programmes in the UK. It is not the final word, nor does it claim to represent the sector's practice, but it is the best indication of programme assessment patterns in UK higher education, based on systematic research conducted through TESTA.

There are three conclusions we draw from the research. The first is that the programme perspective is relatively hidden from view. Our research raises questions about the visibility of different assessment and feedback patterns to those who assure the quality of degrees. From a modular perspective, students appear to have comparable summative assessment loads (often assured through word count equivalence) but from a programme perspective, assessment demands differ markedly. From a teaching perspective, programme data casts doubt about whether lecturers know how much summative, how little formative, and how many varieties of assessment students typically encounter on a whole programme. These hidden dimensions of programme assessment lead us to conclude that academics assess students in modular silos. At worst, modules are being planned and taught in isolation. At best, there is a patchy understanding of the programme hinterland, as academics encounter structural barriers to understanding the balance and range of assessment practice carried out across the programme.

The second conclusion arises from the first: that is, that seeing assessment through the lens of the module does not accurately reflect the lived assessment experience of students. A modular view reflects the academic experience and masks the reality that students are enrolled on three or four concurrent modules; they experience the connections and disconnections of assessment tasks; they delete learning from modules no longer deemed relevant; they juggle multiple deadlines across modules; and they wait for feedback which may or may not help them to improve the next task. Students encounter a programme of study, not a single module. Their programme of study may be single honours or combined honours; it may have more or less coherence and connection, but from a student perspective, it is a learning experience which ranges beyond the single module.

The final conclusion we draw is that the modular structure has accentuated atomised approaches to designing a programme. It has enabled academics to prioritise the value of 'my' module over 'our' programme. Evidence in this study, both in wide variations and in common patterns, suggests that programme design has become an individualistic and modular affair, compounding critiques about its under-theorised nature (Barnett and Coate 2004). This is not surprising. Technical-rational approaches to programme design abound in university quality assurance processes, and are well rewarded. The successful programme contains well-written and tidy documents, clear boxes, neatly expressed learning outcomes, and plausible summative assessment with enough variety to be described as innovative. The problem is not that they are technically sound. The problem is that they do not necessarily represent a complex and creative process of discussion and design, leading to a shared programme approach (Knight 2001). TESTA's audit data is indicative of accidental design, played out within the default structures of the modular system and the technical-rational environment of quality assurance.

In working with many programmes across the sector on the change process linked to participating in the TESTA process, there are signs of hope. Discussing TESTA data with programme teams goes some way to addressing our first conclusion, in bringing to light the hidden and less visible features of

assessment and feedback across a whole programme. Secondly, discussion over TESTA data engenders a shift in academics' perspectives: from a modular view focused on teaching delivery to a programme view focused on the lived assessment experience of students. Finally, the TESTA approach is being used in several universities to bring evidence and team discussion into the curriculum design process through periodic review. This fosters the discussion, dialogue and connected assessment design which is at the heart of a social practice theory of curriculum formation in the 21st century (Knight 2001; Shay 2011; Blackmore and Kandiko 2012).

References

Barnett, R. and Coate, K. 2004. *Engaging the curriculum in higher education*. Maidenhead: Society for Research into Higher Education & Open University Press.

Becher, T., and P. Trowler. 2001. *Academic Tribes and Territories: Intellectual Enquiry and the Cultures of Disciplines*. Buckingham: Open University Press.

Becker, H.S., Geer, B. and Hughes, E. 1968. *Making the grade: Academic side of college life*. New York: Wiley.

Biggs, J.B. and Tang, C. (2011) *Teaching for quality learning at university: What the student does*. 4th edn. Maidenhead: Society for Research into Higher Education & Open University Press.

Biglan, A. 1973a. "The Characteristics of Subject Matter in Different Academic Areas." *Journal of Applied Psychology* 57(3) 195–203.

Biglan, A. 1973b. "Relationships between Subject Matter Characteristics and the Structure and Output of University Departments." *Journal of Applied Psychology* 57(3) 204–13.

Black, P. and Wiliam, D. 1998. "Assessment and classroom learning." *Assessment in Education: Principles, Policy & Practice* 5(1) 7-74. doi: 10.1080/0969595980050102.

Blackmore, P. and Kandiko, C.B. eds. 2012. *Strategic curriculum change: Global trends in universities*. London and New York: Routledge and Society for Research into Higher Education.

Bloxham, S., and P. Boyd. 2007. "Planning a Programme Assessment Strategy." Chap. 11 in *Developing Effective Assessment in Higher Education*. 157–175. Maidenhead: Open University Press.

Boud, D. 2000. "Sustainable Assessment: Rethinking assessment for the learning society." *Studies in Continuing Education* 22(2) 151-167.

Boud, D. and Falchikov, N. 2006. "Aligning assessment with long-term learning." *Assessment & Evaluation in Higher Education* 31(4), 399–413. doi: 10.1080/02602930600679050.

Boud, D. and Molloy, E. 2013. "Rethinking models of feedback for learning: The challenge of design." *Assessment & Evaluation in Higher Education*. 38(6) 698–712. doi: 10.1080/02602938.2012.691462.

Calvo, R. A., Howard, and Markauskaite, L. 2010. "Factors affecting students' experiences and satisfaction about teaching quality in engineering." *Australasian Journal of Engineering Education*. 16(2), 139-148.

Carless, D. 2007. "Learning-oriented assessment: Conceptual bases and practical implications." *Innovations in Education and Teaching International*, 44(1) 57–66. doi: 10.1080/14703290601081332.

Carless, D. 2015. *Excellence in university assessment: Learning from award-winning practice*. Abingdon: Routledge.

Gibbs, G. 2013. "Teaching intelligence - it is possible to avoid the negative mass effects." *Times Higher Education*. 17 January. Accessed 22 May 2016.
<https://www.timeshighereducation.com/teaching-intelligence-it-is-possible-to-avoid-the-negative-mass-effects/422387.article>

Gibbs, G. and Dunbar-Goddet, H. 2009. "Characterising programme-level assessment environments that support learning." *Assessment & Evaluation in Higher Education* 34(4) 481–489. doi: 10.1080/02602930802071114.

Gibbs, G. and Simpson, C. 2004. "Conditions under which assessment supports students' learning". *Learning and Teaching in Higher Education*. 1(1) 3-31.

Hattie, J. 2007. The Power of Feedback. *Review of Educational Research*. 77(1) 81-112.

Harland, T., McLean, A., Wass, R., Miller, E. and Sim, K. N. 2015. "An assessment arms race and its fallout: High-stakes grading and the case for slow scholarship." *Assessment & Evaluation in Higher Education* 40(4) 528–541. doi: 10.1080/02602938.2014.931927.

Hounsell, D. and Anderson, C. 2009. "Ways of Thinking and Practising in Biology and History: Disciplinary Aspects of Teaching and Learning Environments." Chapter 6 in Kreber, C. (ed.) *The university and its disciplines: Teaching and learning within and beyond disciplinary boundaries*. 71-83. New York: Taylor & Francis.

Jessop, T, McNab, N and Gubby, L. 2012. Mind the gap: An analysis of how quality assurance processes influence programme assessment patterns. *Active Learning in Higher Education*. 13(3). 143-154.

Jessop, T., Hakim, Y. El and Gibbs, G. 2014. "The whole is greater than the sum of its parts: A large-scale study of students' learning in response to different programme assessment patterns." *Assessment & Evaluation in Higher Education* 39(1) 73-88. doi: 10.1080/02602938.2013.792108.

Jessop, T. and Maleckar, B. 2014. "The influence of disciplinary assessment patterns on student learning: A comparative study." *Studies in Higher Education* 41:4, 696-711, DOI: 10.1080/03075079.2014.943170

Knight, P. T. 2001. "Complexity and Curriculum: A Process Approach to Curriculum-making." *Teaching in Higher Education* 6 (3): 369–381.

Knight, P.T. 2002. "The Achilles' heel of quality: The assessment of student learning." *Quality in Higher Education*. 8(1) 107–115. doi: 10.1080/13538320220127506.

Knight, P. T. and Yorke, M. 2003. *Assessment Learning and Employability*. Buckingham: Open University Press.

Kuh, G. 2003. "What we're learning about student engagement from NSSE: Benchmarks for Effective Educational Practice." *Change* 35(28) 24-32.

Lizzio, A., Wilson, K. and Simons, R. 2002. "University students' perceptions of the learning environment and academic outcomes: Implications for theory and practice." *Studies in Higher Education*. 27(1) 27–52. doi: 10.1080/03075070120099359.

Nicol, D. J. and Macfarlane-Dick, D. 2006. "Formative assessment and self-regulated learning: A model and seven principles of good feedback practice." *Studies in Higher Education* 31(2) 199–218. doi: 10.1080/03075070600572090.

Price, M., Handley, K., Millar, J. and O'Donovan, B. 2010. "Feedback: All that effort, but what is the effect?" *Assessment & Evaluation in Higher Education* 35(3) 277–289. doi: 10.1080/02602930903541007.

Ramsden, P. 2003. *Learning to teach in higher education*. 2nd edn. New York: Taylor & Francis.

Sadler, D.R. 1989. "Formative assessment and the design of instructional systems." *Instructional Science* 18, 119-144.

Sadler, D.R. 2007. "Perils in the meticulous specification of goals and assessment criteria." *Assessment in Education: Principles, Policy & Practice* 14(3) 387–392. doi: 10.1080/09695940701592097.

Shay, S. 2011. "Curriculum formation: A case study from history." *Studies in Higher Education* 36(3) 315–329. doi: 10.1080/03075071003706681.

Shulman, L. S. 2005. "Signature pedagogies in the professions" *Daedalus* 134(3) 52–59. doi: 10.1162/0011526054622015.

TESTA 2010-16. Accessed 28 January 2016. <http://www.testa.ac.uk>

Torrance, H. 2007. "Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning." *Assessment in Education: Principles, Policy & Practice* 14(3), 281–294. doi: 10.1080/09695940701591867.

Weaver, M.R. 2006. "Do students value feedback? Student perceptions of tutors' written responses", *Assessment & Evaluation in Higher Education*, 31(3), pp. 379–394. doi: 10.1080/02602930500353061.

Williams, J. and Kane, D. 2009. "Assessment and feedback: Institutional experiences of student feedback, 1996 to 2007." *Higher Education Quarterly* 63(3) 264–286. doi: 10.1111/j.1468-2273.2009.00430.x.

Warren Piper, D., D. D. Nulty, and G. O'Grady. 1996. "Examination Practices and Procedures in Australian Universities." Higher Education Division, Department of Employment Education & Training. Evaluations and Investigations Program. AGPS, Canberra.